

ABSTRACT

A method of grammar learning from a corpus comprises, for the other non-context words, generating frequency vectors for each non-context token in a corpus based upon counted occurrences of a predetermined relationship of the non-context tokens to identified context tokens. Clusters are grown from the frequency vectors according to a lexical correlation among the non-context tokens.